2022 IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS 2022)

Singapore 22-24 May 2022



IEEE Catalog Number: CISBN: 9'

CFP22PER-POD 978-1-6654-5955-6

Copyright © 2022 by the Institute of Electrical and Electronics Engineers, Inc. All Rights Reserved

Copyright and Reprint Permissions: Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limit of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

For other copying, reprint or republication permission, write to IEEE Copyrights Manager, IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08854. All rights reserved.

*** This is a print representation of what appears in the IEEE Digital Library. Some format issues inherent in the e-media version may also appear in this print version.

 IEEE Catalog Number:
 CFP22PER-POD

 ISBN (Print-On-Demand):
 978-1-6654-5955-6

 ISBN (Online):
 978-1-6654-5954-9

Additional Copies of This Publication Are Available From:

Curran Associates, Inc 57 Morehouse Lane Red Hook, NY 12571 USA Phone: (845) 758-0400

Fax: (845) 758-2633

E-mail: curran@proceedings.com Web: www.proceedings.com



2022 IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS)

ISPASS 2022

Table of Contents

Message from the General Chair xi
Message from the Program Chair xiii
Organizing Committeexiv
Program Committee xvi
Steering Committee xviii
Keynotes xix
Sponsors
Paper Session 1: Best Paper Nominations
Profiling Intel Graphics Architecture with Long Instruction Traces
Performance Analysis and Optimization with Little's Law
The Indigo Program-Verification Microbenchmark Suite of Irregular Parallel Code Patterns
gpuFI-4: A Microarchitecture-Level Framework for Assessing the Cross-Layer Resilience of Nvidia GPUs
Paper Session 2: Simulation Technologies and Performance Analysis
Distilling the Real Cost of Production Garbage Collectors

Scale-Model Architectural Simulation
MEGsim: A Novel Methodology for Efficient Simulation of Graphics Workloads in GPUs
MARTA: Multi-Configuration Assembly pRofiler and Toolkit for Performance Analysis
Left-Shifter: A pre-Silicon Framework for Usage Model Based Performance Verification of the PCIe Interface in Server Processor System on Chips
Paper Session 3: Code Generation and Optimization
Fourst: A Code Generator for FFT-Based Fast Stencil Computations
Flexible Binary Instrumentation Framework to Profile Code Running on Intel GPUs
POSET-RL: Phase Ordering for Optimizing Size and Execution Time Using Reinforcement

Poster Session 1

(Featur: Hardware Feature Extraction for DNN Auto-Tuning Jorge Sierra Acosta (Universitat Politècnica de Catalunya, Spain), Andreas Diavastos (Universitat Politècnica de Catalunya, Spain), and Antonio Gonzalez (Universitat Politècnica de Catalunya, Spain)	132
GXGauge: A Comprehensive Benchmark Suite for Intel SGX Sandeep Kumar (Indian Institute of Technology, Delhi, India), Abhisek Panda (Indian Institute of Technology, Delhi, India), and Smruti R. Sarangi (Indian Institute of Technology, Delhi, India)	135
SAPCo Sort: Optimizing Degree-Ordering for Power-Law Graphs Mohsen Koohi Esfahani (Queen's University Belfast, UK), Peter Kilpatrick (Queen's University Belfast, UK), and Hans Vandierendonck (Queen's University Belfast, UK)	138
TILE-SIM: A Systematic Approach to Systolic Array-Based Accelerator Evaluation	141
FASE: A Fast, Accurate and Seamless Emulator for Custom Numerical Formats	144
Meterstick: Benchmarking Performance Variability in Cloud and Self-Hosted Minecraft-Like	
Games Jerrit Eickhoff (Delft University of Technology, Netherlands), Jesse Donkervliet (Vrije Universitiet Amsterdam, Netherlands), and Alexandru Iosup (Vrije Universitiet Amsterdam, Netherlands)	147
Simulating Noisy Channels in DNA Storage Mayank Keoliya (National University of Singapore, Singapore), Puru Sharma (National University of Singapore, Singapore), and Djordje Jevdjic (National University of Singapore, Singapore)	150
OS-Level Implications of Using DRAM Caches in Memory Disaggregation Bin Gao (National University of Singapore, Singapore), Hao Wei Tee (National University of Singapore, Singapore), Alireza Sanaee (Queen Mary University of London, United Kindom), Boon Jun Soh (National University of Singapore, Singapore), and Djordje Jevdjic (National University of Singapore, Singapore)	153
VIPP: Validation-Included Precision-Parametric N-Body Benchmark Suite Shigeyuki Sato (The University of Tokyo, Japan), Kota Iizuka (Fixstars Corp., Japan), Naoki Yoshifuji (Fixstars Corp., Japan), and Masaki Natsume (Fixstars Corp., Japan)	156

High-Performance Deployment of Text Detection Model: Compression and Hardware Platform Considerations
Nupur Sumeet (Tata Consultancy Services Research, India), Karan Rawat (Tata Consultancy Services Research, India), and Manoj Nambiar (Tata Consultancy Services Research, India)
Paper Session 4: Benchmarks and Characterization of Emerging Workloads, Scheduling and Reliability
Roofline Model for UAVs: A Bottleneck Analysis Tool for Onboard Compute Characterization of Autonomous Unmanned Aerial Vehicles
RTRBench: A Benchmark Suite for Real-Time Robotics
Characterization of MPC-Based Private Inference for Transformer-Based Models
Spatiotemporal Strategies for Long-Term FPGA Resource Management
Eris: Fault Injection and Tracking Framework for Reliability Analysis of Open-Source Hardware
Poster Session 2
Understanding Data Compression in Warehouse-Scale Datacenter Services

LoopIn: A Loop-Based Simulation Sampling Mechanism	224
Building a Performance Model for Deep Learning Recommendation Model Training on GPUs 2 Zhongyi Lin (University of California, Davis, USA), Louis Feng (Meta Platforms, Inc, USA), Ehsan K. Ardestani (Meta Platforms, Inc, USA), Jaewon Lee (Meta Platforms, Inc, USA), John Lundell (Meta Platforms, Inc, USA), Changkyu Kim (Meta Platforms, Inc, USA), Arun Kejariwal (Meta Platforms, Inc, USA), and John D. Owens (University of California, Davis, USA)	227
Advancing Near-Data Processing with Precise Exceptions and Efficient Data Fetching	230
Profiling an Architectural Simulator	233
Benchmarking Test-Time Unsupervised Deep Neural Network Adaptation on Edge Devices 2 Kshitij Bhardwaj (Lawrence Livermore National Laboratory, USA), James Diffenderfer (Lawrence Livermore National Laboratory, USA), Bhavya Kailkhura (Lawrence Livermore National Laboratory, USA), and Maya Gokhale (Lawrence Livermore National Laboratory, USA)	236
GPUCalorie: Floorplan Estimation for GPU Thermal Evaluation	239
ARBench: Augmented Reality Benchmark For Mobile Devices	242
Cross-Level Characterization of Program Behavior	245
A SIMT Analyzer for Multi-Threaded CPU Applications	248

Microarchitectural Performance Evaluation of AV1 Video Encoding Workloads
Paper Session 5: Machine Learning Systems, Acceleration and Applications
Ruby: Improving Hardware Efficiency for Tensor Algebra Accelerators Through Imperfect Factorization
Pareto Rank Surrogate Model for Hardware-Aware Neural Architecture Search
Learning A Continuous and Reconstructible Latent Space for Hardware Accelerator Design
Bifrost: End-to-End Evaluation and Optimization of Reconfigurable DNN Accelerators
Paper Session 6: Memory Technologies and Architectures
PCMCsim: An Accurate Phase-Change Memory Controller Simulator and its Performance Analysis Hyokeun Lee (Seoul National University, South Korea), Hyungsuk Kim (Seoul National University, South Korea), Seokbo Shim (SK Hynix, South Korea), Seungyong Lee (Seoul National University, South Korea), Dosun Hong (SK Hynix, South Korea), Hyuk-Jae Lee (Seoul National University, South Korea), and Hyun Kim (Seoul National University of Science and Technology, South Korea)
Address Translation Conscious Caching and Prefetching for High Performance Cache Hierarchy 311 Vasudha Vasudha (Indian Institute of Technology Kanpur, India) and Biswabandan Panda (Indian Institute of Technology Kanpur, India)
DRAM Bandwidth and Latency Stacks: Visualizing DRAM Bottlenecks
Author Index